



## Classification automatique du langage de données du service hospitalier des urgences

Binbin Xu, Loïck Bourdois, Cédric Gil-Jardine, Eric Tellier, Frantz Thiessard, Marta Avalos-Fernandez, Emmanuel Lagarde

### ► To cite this version:

Binbin Xu, Loïck Bourdois, Cédric Gil-Jardine, Eric Tellier, Frantz Thiessard, et al.. Classification automatique du langage de données du service hospitalier des urgences. 3e Journée Dataquitaire : IA, RO et Data Science, Feb 2020, Talence, France. hal-02615327

**HAL Id: hal-02615327**

**<https://inria.hal.science/hal-02615327>**

Submitted on 22 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# CLASSIFICATION AUTOMATIQUE DU LANGAGE DE DONNÉES DU SERVICE HOSPITALIER DES URGENCES

(Thème 6 : Intelligence Artificielle)

XU, Binbin<sup>1</sup>, BOURDOIS, Loïck (orateur)<sup>1</sup>, GIL-JARDINE, Cédric<sup>1,2</sup>, TELLIER, Eric<sup>1,2</sup>, THIESSARD Frantz<sup>1</sup>, AVALOS-FERNANDEZ Marta<sup>1,3</sup>, LAGARDE Emmanuel<sup>1</sup>

1 Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, UMR 1219, Bordeaux

2 CHU de Bordeaux, Service des Urgences, Bordeaux

3 Equipe SISTM Inria BSO, Talence

Adresses mail : prénom.nom@u-bordeaux.fr

## Résumé :

Des modèles basés sur l'architecture *Transformer* qui intègrent une étape de pré-entraînement non supervisé à objectif prédictif, tels que le GPT-2 (*Generative Pretrained Transformer 2*) ont atteint récemment des succès remarquables. Nous avons adapté et mis en œuvre un modèle de traitement automatique du langage naturel (NLP pour *Natural Language Processing*) permettant de déterminer si un texte libre clinique est de nature traumatique ou non. Nous avons comparé cette approche, nécessitant un nombre d'échantillons annotés réduit, à une approche entièrement supervisée. Nos résultats (basés sur l'AUC et le F1-score) montrent qu'il est possible d'adapter un modèle polyvalent tel que le GPT-2 pour créer un outil puissant de classification de notes de texte libre en français avec seulement un très faible nombre d'échantillons labélisés.

**Mots clés** : Neural Language Model, pre-training, Transformer, GPT-2

## Introduction

Lors de chaque visite aux urgences, le personnel crée une fiche sur le patient comprenant des données catégorielles, telles que le sexe et l'âge, mais aussi du texte libre qui sont les notes cliniques liées à son état (anamnèses). Ce texte est actuellement traité manuellement ce qui implique des coûts en temps de personnel hospitalier très important. L'automatisation du traitement devrait permettre d'une part, la réduction de ce coût et, d'autre part, le développement d'outils de surveillance. C'est l'objectif du projet TARPON (Traitement Automatique des Résumés de Passages aux urgences pour un Observatoire National) en développement à l'INSERM et le Service des urgences du CHU de Bordeaux, par le biais des derniers outils d'apprentissage profond, supervisés et non supervisés, appliqués à l'analyse automatique du langage. Le premier travail qui porte sur l'application du GPT-2 dévoilé en février 2019 par l'équipe d'OpenAI.

## Méthodologie

Deux scénarios sont comparés. Le premier consiste en un entraînement non-supervisé sur 151 930 données non labélisées puis complété avec un entraînement supervisé sur 10 000 données labélisées. Le second consiste en un entraînement supervisé sur 161 930 données labélisées. La comparaison repose sur l'AUC et le F1 score, appliqués sur le même jeu de données test pour les deux scénarios.

## Originalité/Perspective

L'Application d'un modèle récent de NLP à des données textuelles en français des urgences (jargon et abréviations cliniques, inclus) est originale. Tout comme dans les succès montrés dans d'autres domaines, les coûts liés à l'annotation des données sont radicalement réduits tout en garantissant des résultats comparables à du entièrement supervisé.

L'étude des anamnèses mal classées, leur provenance, ainsi que d'optimisations diverses devraient permettre l'amélioration du modèle. Dans une deuxième phase du projet, il est prévu de s'attaquer à une classification multimodale permettant de décrire d'ensemble des mécanismes traumatiques dans l'objectif de construire un observatoire national de traumatologie.

## Références :

- Binbin Xu, Cédric Gil-Jardiné, Frantz Thiessard, Eric Tellier, Marta Avalos, Emmanuel Lagarde.  
*Neural Language Model for Automated Classification of Electronic Medical Records at the Emergency Room. The Significant Benefit of Unsupervised Generative Pre-training.*  
<https://arxiv.org/abs/1909.01136>
- Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever.  
*Language Models are Unsupervised Multitask Learners.*  
<https://d4mucfpksywv.cloudfront.net/better-language-models/language-models.pdf>  
Voir aussi : <https://openai.com/blog/better-language-models/>